



Metoder til måling og evaluering af arbejdsmiljøindsatser - Overblik over evalueringsteori

Baggrundsnotat (marts 2012, version 2) til workshop 1

Nielsen, Klaus T.; Hasle, Peter; Limborg, Hans Jørgen ; Bramming, Pia; Seim, Rikke

Publication date:
2012

[Link back to DTU Orbit](#)

Citation (APA):

Nielsen, K. T., Hasle, P., Limborg, H. J., Bramming, P., & Seim, R. (2012). *Metoder til måling og evaluering af arbejdsmiljøindsatser - Overblik over evalueringsteori: Baggrundsnotat (marts 2012, version 2) til workshop 1*. Center for forskning i virkemidler og arbejdsmiljøindsatser, CAVI.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Metoder til måling og evaluering af arbejdsmiljøindsatser

Overblik over evalueringsteori

Baggrundsnotat (marts 2012, version 2) til workshop 1 udarbejdet af Klaus T. Nielsen, Peter Hasle, Hans Jørgen Limborg, Pia Bramming og Rikke Seim
Center for forskning i virkemidler og arbejdsmiljøindsatser (CAVI)

1. Resumé

Evaluering er en systematisk vurdering af aktiviteter med henblik på at få grundlag for at træffe fremtidige beslutninger. Evaluering af arbejdsmiljøindsatser har fået en fremtrædende rolle i 2020-handlingsplanen, og det er derfor nødvendigt for arbejdsmiljøaktørerne at finde frem til, hvordan evaluering bedst kan foregå på arbejdsmiljøområdet. En eller anden form for evaluering har sådan set altid fundet sted. Fra den tidligste arbejdsmiljøregulering i 1873 har Arbejdstilsynet forsøgt at opføre sine resultater. Fra forberedelsen af arbejdsmiljøloven i starten af 70'erne er omfanget af den forskellige former for evaluering stadig vokset. Denne proces er i høj grad forstærket af handlingsplanerne som startede med "Rent arbejdsmiljø år 2005" og senest er udvidet med 2020-handlingsplanen. Sideløbende med denne indsats har også arbejdsmiljøindsatsen udviklet sig. I starten forholdt man sig til arbejdsmiljøproblemer med klare kausale sammenhænge med helbredet og lige så klare muligheder for at forebygge. Afskærmning af en rundsav er et godt eksempel. Udviklingen de seneste årtier har fremhævet stadig mere komplekse arbejdsmiljøproblemer som handler om det psykiske arbejdsmiljø, muskelskelet besvær, sikkerhed og sygefravær. Det er problemer som både er vanskeligt at beskrive og vanskeligt at løse.

Evalueringsteorien har udviklet sig parallelt med arbejdsmiljøloven. I 60'erne og 70'erne opstod evalueringsteorien, først i USA med det såkaldte 'eksperimentel selskab', hvor der blev udviklet metoder til effektevaluering med stærk inspiration fra den biomedicinske forskning. Som en reaktion på effektevaluering blev der i 80'erne udviklet den responsive evaluering, hvor hovedsigtet er at styrke den indsats, som evalueres gennem forskellige feedback mekanismer. Herefter fulgte New Public Management og monitoreringsbølgen, som forsøger at øge ansvarligheden hos de enkelte aktører gennem løbende monitorering af deres indsats. Op til nutiden, hvor evidens i høj grad er blevet et krav der skal gøre det muligt at prioritere ressourcerne til de samfundsmæssige indsatser. Evidenskravet driver evalueringerne i to retninger. Dels i retning af *effektevaluering* som bygger på den tidlige effektevaluering og prioriterer randomiserede kontrollerede forsøg, dels i retning af *kritisk evaluering* (virkningsevaluering), som prioriterer de virkningsmekanismer, som får noget til at ske. Det kan konkluderes, at det på arbejdsmiljøområdet er nødvendigt med elementer af flere evalueringstilgange. Det skyldes, at evidens på arbejdsmiljøområdet må opfattes som den bedst tilgængelige viden på et givet tidspunkt, fordi der aldrig vil kunne skaffes absolut sikker viden om, hvorvidt komplekse indsatser kan løse komplekse problemer. Omvendt er det nødvendigt at gribe ind overfor de væsentlige arbejdsmiljøproblemer som samfundet har i dag.

2. Hvad er evaluering

Indledningsvist er det vigtigt at præcisere, hvad evaluering er for et fænomen. Evaluering kan defineres som en systematisk bedømmelse af udfald, præstationer og organisering af samfundsmæssige aktiviteter, og det er en bedømmelse som er rettet mod fremtidige handlingssituationer og dermed har evalueringer et anvendelsessigte (Dahler-Larsen 2010; Vedung 1998). Evaluering adskiller sig fra forskning i og med, det har et konkret anvendelsessigte som forskning ikke nødvendigvis har, og evaluering iværksættes af beslutningstagere som ønsker et grundlag for fremtidige beslutninger. Men med disse skillelinjer skal det også fremhæves, at der er et betydeligt sammenfald, hvor evaluering benytter sig af specielt samfundsvidenskabelige metoder og i stigende grad også kvalitetskrav. På samme måde som den samfundsvidenskabelige forskning er evaluering således også præget af store diskussioner og uenigheder både om de overordnede tilgange (paradigmer) for evaluering, og hvilke konkrete metoder, der er bedst egnede. Forskningen har derfor også kastet sig over evaluering som et forskningsobjekt, og evalueringsteori kan give væsentlige bidrag til at forstå, hvad evaluering er, og hvordan og hvornår den bedst kan anvendes.

I dette notat vil vi prøve at rede trådene ud for både, hvorfor evaluering får en stadig større betydning på arbejdsmiljøområdet, og hvordan evaluering som teori og metode har udviklet sig. Vi starter i nutiden og ser på formuleringerne om evaluering, som vi finder dem i den aktuelle 2020-handlingsplan for arbejdsmiljøet og i baggrundsmaterialet for handlingsplanen. Derefter tager vi et historisk vue over evalueringer på arbejdsmiljøområdet over de sidste mange år for at se, hvordan de aktuelle tendenser ser ud i det perspektiv. Indledningsvis er der måske grund til at påpege, at når det gælder den historiske baggrund skal evaluering forstås lidt bredt. Der fremhæves også træk, som ikke i snæver forstand er evaluering, men mere kan siges at udgøre forløbere for de aktuelle evalueringsaktiviteter. Det er klart at arbejdsmiljøområdet langt fra står alene, når det gælder den store interesse for evalueringer, og i den anden halvdel af notatet præsenterer vi derfor udviklingen i evalueringsteori og de aktuelle diskussioner på evalueringsområdet.

3. Hvorfor evaluering af arbejdsmiljøindsatser?

Et første relevant spørgsmål er, hvorfor evaluering af arbejdsmiljøindsatser er blevet et så aktuelt emne. Arbejdsmiljøområdet er jo netop præget af en lang og stærk forskningstradition. Svaret på dette spørgsmål ligger netop i adskillelsen mellem forskning og evaluering. Arbejdsmiljøforskningen interesserer sig for en bred forståelse af hvad arbejdsmiljø er, og hvordan det hænger sammen med helbred og trivsel. En sådan forståelse bygges op over en årrække af mange forskellige forskningsbidrag, mens evaluering interesserer sig for resultater af konkrete indsatser som grundlag for at træffe beslutninger om nye indsatser. Selvom forskningen leverer væsentlig viden som indsatser kan bygge på, kan den sjældent på kort sigt anvendes direkte som grundlag for beslutningstagning. Forskningen vil typisk bidrage med bredere svar med flere nuanceringer, og ofte lang tid efter beslutningstagerne har været nødt til at tage stilling til den næste indsats. Evaluering bliver derfor vigtig for at sikre det bedst mulige grundlag for aktuelle beslutninger.

3.1. 2020-handlingsplanen

Det er denne tankegang som afspejler sig i aftalen mellem regeringen (den daværende – Venstre og Det Konservative Folkeparti), Socialdemokraterne, Dansk Folkeparti og Det Radikale Venstre ”En

strategi for arbejdsmiljøindsatsen frem til 2020" fra marts 2011. Aftalen har en række direkte referencer til overvågning og evaluering som grundlæggende politikredskaber på arbejdsmiljøområdet:

- Politikken er vidensbaseret for som det hedder i den daværende regerings oprindelige udspil i september 2010, så: "... hviler [strategien] på et stærkt fagligt fundament. De seneste år er der indhentet både viden og erfaringer med arbejdsmiljøindsatsen. Arbejdstilsynet, Det Nationale Forskningscenter for Arbejdsmiljø og Arbejdsskadesstyrelsen har lavet grundig analyse og vurdering af de centrale udfordringer ..." (s. 1 i "Nye veje til et bedre arbejdsmiljø – Regeringens strategi for arbejdsmiljøindsatsen frem til 2020", Regeringen). I strategi-papiret henvises da også under overskriften "Et solidt fagligt grundlag" direkte til rapporten "Fremtidens arbejdsmiljø 2020 – Fagligt grundlag for prioritering af arbejdsmiljøindsatsen" som er Arbejdstilsynets rapportering af det indhentede vidensgrundlag.
- Aftalen prioriterer tre områder og fastlægger, hvad den daværende regering i 2010-strategipapiret har kaldt, "Tre ambitiøse og konkrete mål" (s. 4): Kvantificerede reduktionsmål for antallet af arbejdsulykker, antallet af beskæftigede, som er psykisk overbelastede, og antallet af beskæftigede, der har muskelskeletoverbelastninger.
- Strategien, som den kommer til udtryk i aftalen, er bl.a. formuleret som 19 initiativer. Initiativ 18 og 19 har i 2010-strategipapiret fællesoverskriften "Resultaterne skal dokumenteres". Initiativ 18 er "Måling af fremdriften i arbejdsmiljøet" og her konstaterer 2010-teksten, at det i dag ikke er muligt at følge med i, om man kommer nærmere målsætningerne undervejs. Derfor skal der etableres målinger, der kan muliggøre dette. I følge 2011-aftalen skal det endvidere være muligt at følge med i udviklingen på områder der ikke er omfattet af prioriteringen. Initiativ 19 omhandler "effektmålinger af de konkrete indsatser". I 2010-teksten hedder det: "Ved at evaluere effekten af de forskellige aktiviteter får man et reelt grundlag for at sikre, at aktiviteterne bidrager bedst muligt til at opfylde målsætningerne." Og i den endelige aftale: "Der skal stilles krav om, at Arbejdstilsynet og andre arbejdsmiljøaktører, der modtager midler fra det offentlige, skal dokumentere effekten af væsentlige aktiviteter. Målet er at sikre, at aktiviteterne bidrager til at nå målene for arbejdsmiljøindsatsen."
- Men det er ikke kun de sidste to initiativer som trækker på evalueringstankegangen. Initiativ 14, "Bedre vejledning og information om arbejdsmiljø" handler om, at "der iværksættes en ekstern undersøgelse, af hvordan de forskellige arbejdsmiljøaktører kan styrke samarbejdet om at informere og vejlede virksomhederne. ... Undersøgelsen skal også afdække, om den direkte vejledning og information kan organiseres mere hensigtsmæssigt.". Initiativ 10, "Undersøgelse af erfaringerne med virksomhedernes arbejdsmiljøorganisation", stipulerer at en undersøgelse af erfaringerne med omlægningen på området igangsættes før 2013. Og endeligt er der initiativ 1, "Risikobaseret tilsyn – Fokus på virksomheder med arbejdsmiljøproblemer", hvor kernen er, at tilsynet med virksomhederne skal afpasses efter risikoen for, at virksomhederne har problemer med arbejdsmiljøet.

Med hensyn til risikobaseret tilsyn er det nærmest en modebølge indenfor en række reguleringsfelter (Black & Baldwin 2010), som også i høj grad bygger på en evalueringstankegang. Risikobaseret regulering er en fællesbetegnelse, der har stor legitimitet i disse år som Black & Baldwin sammen-

fatter på følgende måde: "Den vigtigste komponent ... er evaluering af risikoen for overtrædelse og en vurdering af den betydning som overtrædelse vil have for reguleringsmyndighedens muligheder for at opnå sin målsætning. I den ideelle form tilbyder risikobaseret regulering en evidensbaseret metode til at målrette brugen af ressourcer og til at prioritere opmærksomhed på den største risiko i overensstemmelse med en gennemskuelig, systematisk og begrundet overordnet ramme." Kernen i risikobaseret regulering adskiller sig således ikke fra centrale dele af evalueringstankegangen, og begrebet risikobaseret tilsyn læner sig derfor også op ad en evalueringstænkning.

Der er således stærke træk i hele den strategi for arbejdsmiljøindsatsen, der gælder for de kommende år som knytter an til aktuelle forestillinger om monitorering, evaluering og evidensbaseret politik. Klarest udtalt i initiativ 19, men også i en række af de andre initiativer. Der refereres til evaluering relativt overordnet og med lidt forskellige ord som både handler om at dokumentere effekter og om at finde ud af, hvordan virksomhederne får størst gavn af indsatser. En afklaring af hvordan evaluering kan anvendes i arbejdsmiljøindsatsen på den bedste måde, er netop formålet med Arbejdsmiljørådets workshoprække og CAVI's bidrag hertil.

3.2. Evaluering af arbejdsmiljøindsatsen i historisk perspektiv

Forventninger om, at regulering og politiklægning på arbejdsmiljøområdet skal ske på et gennemsigtigt, systematisk og forsvarligt grundlag er jo ikke ny, omend forventningerne uden tvivl er taget til i styrke, som vi skal se senere. Denne udvikling skal naturligvis ses i lyset af, at arbejdsmiljøet og vores forståelse af det også udvikler sig. I det følgende vil vi foretage på nogle centrale nedslag i udviklingen af evalueringsbølgen på området.

I en vis forstand ligger det i enhver myndigheds institutionelle natur at opsamle erfaringer og udviklingstræk på kvantitativ vis med henblik at korrigere og begrunde sin egen indsats. Fx anfører Gyldendals Store Danske Encyklopædi: "Ordet statistik kommer via ty. fra nylat. statisticus 'hvad der vedrører statssager'" og "Statistik, videnskab, som i sin oprindelige betydning omfatter en metodisk, empirisk funderet beskrivelse af samfundsforhold vha. indsamling og systematisering af talmaterialer." Danmarks Statistik blev etableret i 1850, og stat og statistik er sammenbundet fra første færd i etableringen af den grundlovsbaserede forvaltning i Danmark. I Arbejdstilsynets første årsberetning, "Beretning om Arbejdstilsynets Virksomhed indtil Udgangen af Aaret 1874", kan vi fx finde følgende tabel over børnearbejdere på fabrikker:

Efter de fremkomne Oplysninger var Bestanden af Fabrikkerne den 1ste Decbr. 1873 følgende:

Bedrifternes Art.	Antal Bedrifter inddragne.	B o r n				Unge Mennesker 14—18 Aar.	
		under 10 Aar.		fra 10—14 Aar.			
		Mandskøn.	Kvindeskøn.	Mandskøn.	Kvindeskøn.	Mandskøn.	Kvindeskøn.
Bernstøberier og Ma- skinfabrikker	67	"	"	6	"	389	"
Mekaniske Fabrikker etc.	42	"	"	3	"	113	"
Andre Metalvarefabrikker	13	6	1	14	"	23	2
Papirfabrikker	5	"	"	2	3	5	7
Forskjellige Papir- og Pappvarefabrikker	9	5	7	23	16	12	25
Tæpet-, Kulslegardin- og Børdugfabrikker	10	12	1	60	2	19	1
Pibeslange-, Børstefab- rikker o. desl.	3	"	1	5	39	"	13
Lædvare- og Skorfab- rikker	6	3	"	3	"	9	"
Bogtrykkerier og Sten- trykkerier	49	"	"	49	5	156	26
Uldvarefabrikker	64	8	3	123	78	113	54
Bomulds- og Linned- varefabrikker	19	12	6	98	56	76	93
Forskjellige kemiske Fa- brikker	27	1	3	59	17	41	21
Tændstikfabrikker	7	12	"	93	48	27	28
Cement-, Sten- og Stukfaturfabrikker	3	"	"	7	2	3	2
Porcellæn-, Faience- og Lervarefabrikker	18	28	"	74	10	88	6
Glasværker	5	10	"	66	3	44	"
Tobak- og Cigarfabrikker	116	282	26	906	173	387	88
Dampmøller	8	"	"	"	"	27	"
Chokolade- og Sutter- varefabrikker	17	"	"	12	3	16	23
Ekstorfabrikker	6	5	2	48	19	21	1
	494	384	50	1651	474	1569	380
		434		2125		1949	
		2559					

Figur 1: uddrag fra kopi fundet på nettet (<http://www.emu.dk/gym/fag/hi/inspiration/tema/bornarb/28tabel1874.pdf>)

Vi kan henføre vores gældende arbejdsmiljølovgivning til Arbejdsmiljøloven af 1975 – som jo siden kun er blevet ændret gennem mindre omfattende lovændringer. Da arbejdsmiljøloven blev udarbejdet var det blandt andet på grundlag af fire rapporter fra Arbejdsmiljøgruppen af 1972 (1973; 1974; 1975; 1976), som langt hen af vejen kan opfattes som en systematisk evaluering af arbejdsmiljøtilstanden i begyndelsen af 1970'erne.

Med handlingsprogrammet "Rent arbejdsmiljø år 2005" fra 1996 begyndte Arbejdstilsynet at formulere målbare prioriteringer for sig selv og det samlede arbejdsmiljøsystem. I 1996 var de syv visioner dog ikke mere målsatte end at de udpegede problemer bare skulle reduceres mest muligt/helt undgå, som det hed.

Ikke kun arbejdsmiljøsystemet selv stod som initiativtager til den forøgede fokus på evaluering. Rigsrevisionen spillede en væsentlig rolle i form af en undersøgelse af Arbejdstilsynet i 2000 af Arbejdstilsynets styring af tilsynsfunktionen (Rigsrevisionen 2000). Nogle af de centrale/illustrative punkter i sammenfatningen var, at undersøgelsen viste, at:

- "uanset at Handlingsprogram 2005 har skullet virke siden 1996, udviser antallet af arbejdsulykker og -lidelser, opdelt på de 7 visionsområder og opgjort på baggrund af de til Arbejdstilsynet anmeldte arbejdsskader, ikke en klar stigende eller faldende tendens. (...),"
- "Arbejdstilsynets resultatkontrakt ikke indeholder mål for og resultatkrav til kvalitet og produktivitet samt effekt af Arbejdstilsynets opgavevaretagelse,"
- "Arbejdstilsynet i undersøgelsesperioden ikke har foretaget produktivitetmålinger ud over opgørelse af enhedsomkostningerne for udvalgte aktiviteter,"
- "Arbejdstilsynet ikke har et samlet ledelsesinformationssystem, der systematisk opgør resultater, kvaliteten heraf, produktivitet og effekt".

Der var således et betydeligt pres på en forøget evalueringsvirksomhed, og det var måske en af årsagerne til, at der i 2002 blev strammet op på 2005 handlingsplanen med konkrete reduktionsmål for nogle af visionerne (alvorlige ulykker – 15 %, tunge løft – 15 %, ensidigt, gentaget arbejde – 10 %, psykisk arbejdsmiljø – 5 %). En vigtig detalje i "Rent arbejdsmiljø år 2005" var også ambitionen om "at udbygge statistik og dokumentation både i Arbejdstilsynet og Arbejdsskadestyrelsen så alle arbejdsmiljøaktører – herunder virksomheder – kan bruge det statistiske materiale." Der har siden været en væsentlig forøget monitorering af arbejdsmiljøet, fx har man udarbejdet de såkaldte overvågningsrapporter siden 1999 og gennemført de såkaldte VOV-undersøgelser (VOV for VirksomhedsOverVågning) siden 2001.

Handlingsprogrammet "Rent arbejdsmiljø 2005" blev evalueret i rapporten "10 år med handlingsprogrammet Rent Arbejdsmiljø år 2005" (Arbejdstilsynet 2007). Resultatet var lidt nedslående for så vidt som målet ikke blev nået for tre af målsætningerne, mens der for højrepetitivt arbejde blev fundet et fald på 18 %, og dette mål var således blevet opfyldt.

I 2005 kom "Fremtiden arbejdsmiljø 2010" som også havde konkrete reduktionsmål på sine fire prioriterede indsatsområder: Ulykker – 20 %, psykisk arbejdsmiljø – 10 %, muskelskeletbesvær – 10 %, støj – 15 % for høreskadende støj, 10 % for øvrig støj. En særlig detalje er, at reduktionsmålene denne gang var formuleret som effekt på sygefraværet – ikke som reduktion i belastningen. Man kan sige, at evalueringsteknologierne sofistikeres, og målsætningerne udvikles, så de er mere direkte rettet mod det man egentligt ønsker at opnå, da der var stærk fokus på arbejdsudbuddet i denne periode.

I forbindelse med arbejdsmiljøreformen 2004 iværksatte Arbejdstilsynet en større evaluering af reformen. Evalueringen er afrapporteret i en række arbejdsrapporter, fire delrapporter og en slutrap-

port (Arbejdstilsynet 2010). En lang række aktører eller institutioner er også blevet evalueret i de senere år, fx:

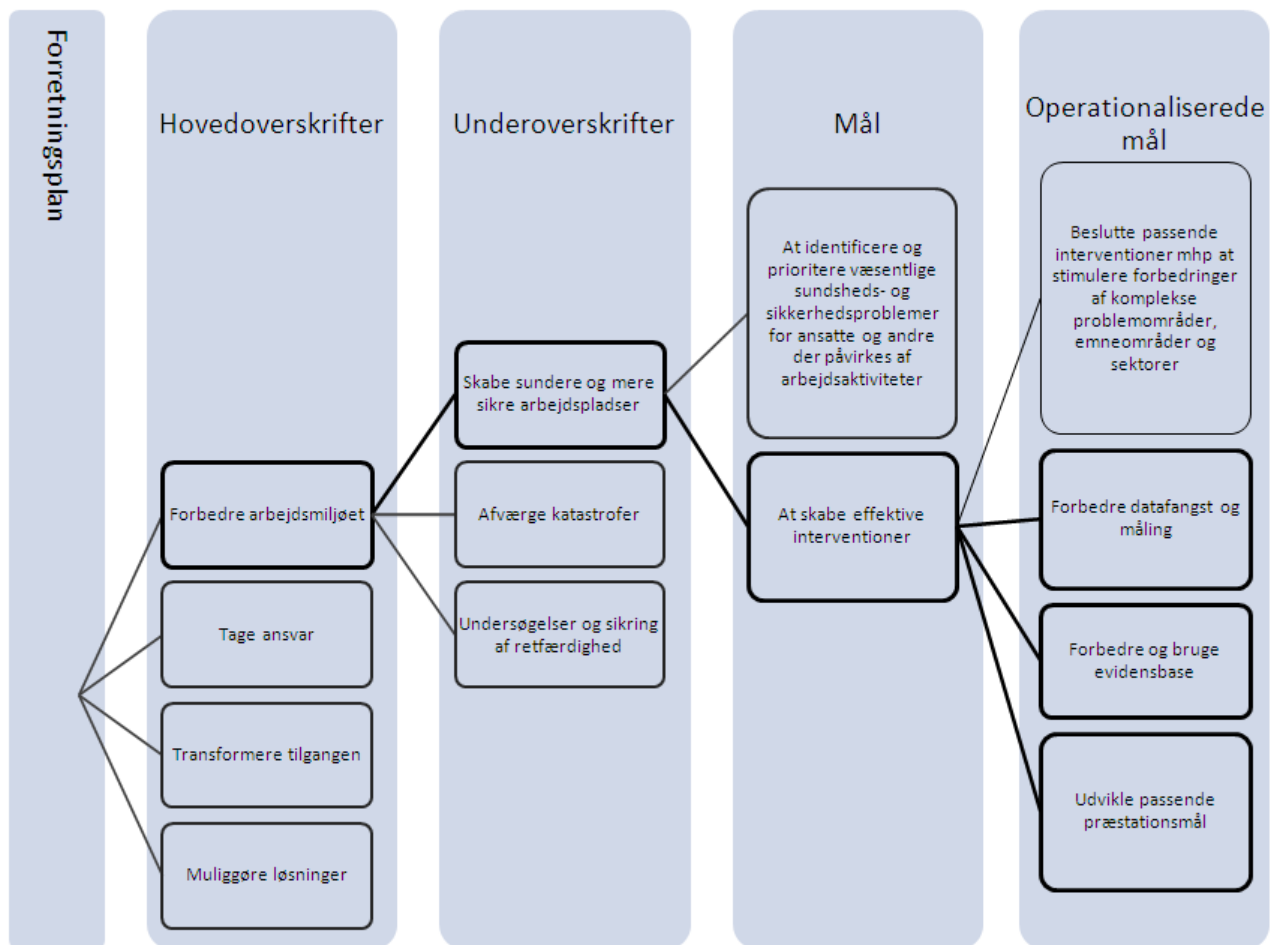
- "Evaluering af Videncenter for Arbejds miljø – Centerets virksomhed 2005 til primo 2008" (Advice A/S 2008)
- "Stikprøvevis effektevaluering af branchearbejds miljørådenes projekter" (Capacent & Arbejdstilsynet 2007)
- Evaluering af Arbejds miljøforskningsfonden (Arbejdstilsynet 2011)

Således er evalueringer på arbejds miljøområdet ikke nogen ny foreteelse, men har spillet en væsentlig rolle for såvel den faglige som den politiske udvikling på området i hvert fald siden årtusindeskiftet, men med lange rødder tilbage i historien.

Man kan finde parallelle træk i udviklingen i andre landes arbejds miljøstrategier. Vi viser i nedenstående boks et enkelt, udvalgt, illustrativt eksempel fra det britiske Arbejdstilsyn (Health & Safety Executive (HSE)). HSE udarbejder årlige såkaldte *business plans* (forretningsplaner), og i planen for 2010/11 (Health and Safety Executive 2012) ses for perioden 2000-10 fundet følgende tre mål om reduktion af helbredsproblemer:

- Antal tabte arbejdsdage grundet arbejdsrelaterede ulykker og sygdom pr 100.000 arbejdere med 30 %
- Forekomsten af arbejdsrelateret sygdom med 20 %
- Forekomsten af dødsfald og alvorlige ulykker med 10 %

Sådanne overordnede mål kan ikke direkte omsættes til konkrete aktiviteter. Derfor er forretningsplanen bygget op om et målhierarki. Vi har i nedenstående figur 2 udvalgt, og dermed fremhævet, særlige dele af målhierarkiet med henblik på at vise hvordan målinger og evidens anvendes til at omsætte målene til et konkret operationaliserbart niveau.



Figur 2 viser hvordan det overordnede mål 'forbedring af arbejdsmiljøet' er blevet nedbrudt i delmål. Figuren viser kun udvalgte dele af målhierarkiet.

Pointen er, at den udvikling vi har set i Danmark, når det gælder den strategiske tilgang til reguleringen af arbejdsmiljøet, ikke kan ses som isoleret dansk udvikling; men nærmere må ses som et fællestræk i mange lande og i virkeligheden ikke kun på arbejdsmiljø området, men også indenfor en lang række andre reguleringsfelter.

3.3. Manglende viden om virkninger af indsatser

Det er en fuldt forståelig interesse for, hvad der virker, som ligger bag den voksende bølge af initiativer til at evaluere indsatserne på arbejdsmiljøområdet. Det er imidlertid overraskende, hvor ufuldstændig en viden vi faktisk har, når det gælder virkningen af forskellige reguleringer og indsatser. I et omfattende litteraturstudie, som bl.a. har fokus på effekten af reguleringen, finder Hansen et al. (2009), at en række studier af konkrete arbejdsmiljøindsatser tilsyneladende kun har lille eller ingen effekt. Omvendt findes der også en række studier, som kan tages til indtægt for at arbejdsmiljøregulering og -indsatser har en effekt. Det er karakteristisk, at undersøgelserne alle har relativt snævre afgrænsninger. De fleste af undersøgelserne ser på tilsyn, ofte bødeafgivelse, og effekten på arbejdsskader - primært arbejdsulykker. Hansen et al. kigger også på undersøgelser, der vurderer forskellige tilsynsstrategier. De konkluderer på den ene side, at der er indikationer på at en formel og problemspecificerende strategi skulle være mest økonomisk fordelagtig, og på den anden side, at der er en række undersøgelser, der påviser effekter af mere samarbejdsorienterede og rådgivende strategier (s. 37).

Der er lavet en større britisk litteraturanalyse på feltet: "What works in delivering improved health and safety outcome – A review of the existing evidence" (Cox et al. 2008) som peger på en række af de områder, hvor indsatser ser ud til at virke. Grundlaget er 846 HSE research reports (dvs. primært rapporter, der er bestilt og publiceret i HSE's egen rapport serie) og 591 videnskabelige artikler fundet ved søgninger i videnskabelige databaser. Efter en nærmere vurdering har de fundet 66 publikationer fra 2002 eller senere med relevans for spørgsmålet "What works?". Ud over de reviewede artikler har man også interviewet 31 personer i det britiske arbejdstilsyn til brug for rapporten. Rapporten bygger videre på en rapport fra 2001, og i sammenligning hermed konkluderer Cox et al.: "Vores vurdering af kvaliteten af evidens er, at den er blandet. Evidensen er i nogle nyere indsatsområder som fx bedriftssundhedstjeneste og metoder til at forebygge MSB, relativt godt designet og robust. Men en væsentlig del af evidensgrundlaget fortsætter med at lide af vedvarende svagheder med små, om nogen, forbedringer siden litteraturgennemgangen blev gennemført i 2001." (p. 62).

Cox et al. har en række mere specifikke konklusioner, som kan findes i nedenstående tekstboks. På den ene side er de relevante for prioritering af indsatser, men på den anden side kan de være lidt vanskelige at vurdere ude af deres oprindelige sammenhæng.

Centrale fund angående 'Hvad virker' i forhold til helbreds- og sikkerhed:

- **Tilsyn og opmærksomhedsøgende ansigt-til-ansigt begivenheder** spiller en vigtig rolle i storstilede, flerstrengede kampagneaktiviteter. Partnere og andre aktører kan spille en vigtig rolle. Skriftlige input kan i nogen grad være effektive til at forøge opmærksomheden i forhold til risici, men er ikke effektive ift. at ændre opmærksomhed til adfærdsændring. Flerstrengede kampagner synes at være mere effektive, da de viser evidens i forhold til adfærdsændringer hos en mindre andel af arbejdsgiverne.
- **Nye interventioner virker.** Bedriftssundhedstjenestens indsatser overfor muskelskeletbesvær fører til forbedret sundhed og sikkerhed.
- **Regulering** skaber forbedret sundheds- og sikkerhedspraksisser blandt dem, der er villige til at efterleve reguleringen, men 'usynlige' risici som fx arbejdsrelateret stress er sværere at føre tilsyn med og er derfor mindre egnet til regulering.
- **Værktøjs- og sikkerheds- og sundhedsdage** er virkningsfulde over for villige og engagerede medarbejdere.
- **Medarbejderinvolveringsteknikker** såsom arbejdsmiljø- og sikkerhedsrepræsentanter kan være effektive til udbredelse af sundheds- og sikkerhedsinformation og til at skabe adfærdsændringer.
- **Målrettede initiativer** kan være hjælpsomme til at fastsætte målsætninger og til at opmuntre positive tiltag i brancherelaterede sundheds- og sikkerhedsforbedringer ved at skabe engagement hos brancheorganisationer.

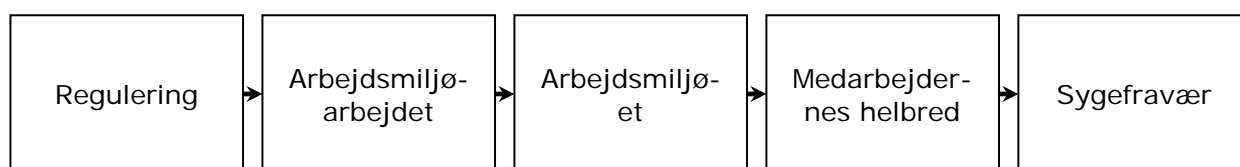
Cox et al. (2008), s. vii

Denne korte gennemgang af forskningslitteraturen på feltet peger på, at det ikke er muligt at drage håndfaste konklusioner i forhold til hvilke strategier for arbejdsmiljøregulering og -indsatser, som er mest effektive.

4. Hvorfor er det svært at evaluere arbejdsmiljøindsatser?

Arbejdsmiljøindsatser er jo ikke en ny foreteelse. Og behovet for at dokumentere betydningen af indsatsen og for at kunne prioritere forskellige reguleringsindsatser på et solidt fagligt grundlag er heller ikke af ny dato. Hvorfor er vores viden på feltet så, så beskeden som det åbenbart er tilfældet?

Svaret skal søges i, at der er en lang virkningskæde mellem den indsats man gør i arbejdsmiljøsystemet, og det man ønsker at opnå i den sidste ende. Denne kæde er skildret i nedenstående figur.



Figur 3 illustrerer virkningskæden på arbejdsmiljøområdet.

Der er imidlertid ikke klare årsags-virknings-sammenhænge mellem de enkelte led i kæden. I praksis er det derfor svært eller måske umuligt at opnå en "alt andet lige"-situation, hvor en relevant kvantificering af effekten længere nede af virkningskæden kan lade sig gøre. Der er flere forhold som udfordrer tankegangen om direkte årsags-virknings-sammenhænge.

I den oprindelige arbejdsmiljøregulering, før den nuværende arbejdsmiljølov, bestod kæden kun af tre led: Regulering – arbejdsmiljø – helbred. Her stillede myndighederne klare krav, fx om afskærmning på rundsave som efterfølgende blev kontrolleret. I et sådant tilfælde kunne der relativt enkelt konstateres en sammenhæng mellem afskærne fingre og tilstedeværelsen af afskærmning i træindustrien. Sådanne enkle sammenhænge genfindes imidlertid kun i en mindre del af arbejdsmiljøet, hvilket netop var en væsentlig årsag til, at en arbejdsmiljølov med et universelt arbejdsgiveransvar og refleksiv evaluering blev indført. Det blev således nødvendigt at indføre arbejdsmiljøarbejdet som en mellemvariabel. Opmærksomheden er i stigende grad blevet rettet mod netop arbejdsmiljøarbejdet med diskussioner af sikkerhedsorganisationen i 1980'erne (Møller, Jensen, & Broberg 1988; Stranddorf, Møller, & Jensen 1992) og i det seneste årti systematisk arbejdsmiljøledelse (Frick & Wren 2000; Hasle & Zwetsloot 2011). Dette ekstra led i kæden er på den ene side nødvendigt for at kunne påvirke det stadig mere komplekse arbejdsmiljø, og på den anden side er arbejdsmiljøarbejdet i virksomhederne en mangesidig størrelse, som vanskeligt kan indfanges i simple modeller.

Hvis pilene mellem de forskellige led i virkningskæden var stærke kausalitetsmekanismer, ville sammenknytningen mellem kædens yderste led også være relativt robust. Men ingen af modellens pile er stærke kausalitetsmekanismer: Ved vi, hvordan forskellige reguleringsstile – typisk benævnt kontrollerende henholdsvis understøttende (Nielsen et al. 2010) – påvirker arbejdsmiljøet fx i små og mellemstore virksomheder? Nej, ikke helt præcist. Ved vi, om et systematisk ledelsesarbejde, som vi fx ser det i certificerede virksomheder, giver et godt arbejdsmiljø? Nej, ikke så nøje (Hasle & Zwetsloot 2011; Robson et al. 2007). Ved vi, hvilke forhold i arbejdsmiljøet, der udløser stressrelateret sygdomme? Jo, men ikke helt præcist (Briner & Reynolds 1999; Semmer 2006). Kender vi koblingen mellem sundhed og sygefravær? Ja, vi ved, at den i hvert fald er mere kompliceret end bare et spørgsmål om, at manglende sundhed giver sygefravær. At etablere viden om virkninger

mellem de forskellige kasser blot to og to er altså langt fra et banalt problem. Og som en konsekvens er vores viden om den samlede længere virkningskæde naturligvis også meget usikker.

4.1. Baggrundsstøj

Men vi er jo ikke uvidende om, at der er en vis sammenhæng mellem de forskellige led i kæden. Og selvom vi ikke har sikker viden om virkningerne hen gennem kæden, betyder det jo ikke, at der ikke er nogen virkning; vi har bare ikke sikre metoder til at skille direkte effekter fra baggrundsstøj.

En væsentlig forudsætning for den måde vores arbejdsmiljøindsats fungerer på er, at mange reguleringsinstrumenter virker i fællesskab i deres forsøg på at ændre adfærden hos virksomhedens ledelse og medarbejdere. Der er informationskampagner fra Videncenter for Arbejdsmiljø, der er arbejdsmiljørådgivere, der er tilsynsførende fra Arbejdstilsynet, der er arbejdsmiljørepræsentanter, der har været på arbejdsmiljøkurser, og der er ledere, der har arbejdsmiljøperformance inkluderet i deres bonusvilkår. Det bliver derfor vanskeligt at adskille et enkeltstående BAR-projekts virkning fra synergien fra de mange andre indsatser, som i princippet søger at opnå til en tilsvarende adfærd i virksomhederne.

Et andet klassisk problem, når det gælder koblingen mellem arbejdsmiljø og sundhed, er latenstiden. Fra udsættelsen for kræftfremkaldende stoffer eller støj til diagnoser som kræft eller varigt høretab går der ofte flere årtier. Andre virkningsmekanismer har væsentligt kortere tid fra påvirkning til effekt, men selv i koblingen mellem arbejdsmiljøarbejde og arbejdsmiljø kan der sagtens være en vis forsinkelse; det er jo netop i det lange seje træk i ledelsen og arbejdsmiljøorganisationen, at grundlaget lægges for en organisation, der kan og vil gøre noget ved arbejdsmiljøproblemerne.

Det måske vanskeligste problem er arbejdsmiljøets øgede kompleksitet. Arbejdsmiljøreguleringen har i stigende omfang fået kompleksitet som et grundvilkår. Der er langt fra børnearbejde til arbejdsrelateret stress. I Arbejdsmiljølovens levetid har arbejdsmiljøproblemstillingerne udviklet sig ganske betydeligt. I den første periode var der en væsentlig udvikling omkring stoffer og materialer, hvor opmærksomheden på kræft, reproduktionsskader, hjerneskader og hudirritation steg betydeligt. Siden er indeklima, muskelskeletbesvær og psykisk arbejdsmiljø blevet tildelt en stadig større opmærksomhed.

Lidt (for) firkantet kan man stille forskellen mellem at regulere høreskadende støj og stress op, som et billede på den stigende kompleksitet der karakteriserer de problemer, som arbejdsmiljøreguleringen skal forholde sig til. Støjen er umiddelbart til at "få øje på"; den kan let måles, hvad der forårsager støjen er ret let at fastslå, at pege på løsninger er ofte rimeligt ligetil (og så måske alligevel ikke, men ...), og virkningen på sundheden har vi et simpelt mål for. Hvis vi skal evaluere indsatsen mod støj i arbejdsmiljøet, er der altså en række værktøjer, som vi kan bruge. Stress-epidemien som nogen kalder det, er tilsyneladende omfattende, men dog ikke så meget, at der ikke hersker stor uenighed om dens realitet og omfang. Der er ingen sikre biologiske markører for stress. Hvad der forårsager stress er ofte helt uklart og åbent for diskussion. Der er kun i begrænset omfang udviklet løsninger som bare i rimeligt omfang kan reducere medarbejdernes stress – for slet ikke at tale, at om sikre at arbejdsforholdene ikke er stressende generelt set – og den langsigtede virkning af stress

på befolkningens sundhed er i bedste fald uklar. Det bliver følgelig vanskeligt at svare på, hvordan man evaluerer arbejdsmiljøindsatser overfor stress?

Endeligt er der også et andet problem med at evaluere en given indsats tilstrækkeligt grundigt til, at man kan sige at have sikker viden om reguleringens virkning; At etablere viden er ikke omkostningsfrit. Det vil typisk være sådan at jo mere præcis viden man ønsker sig, jo længere tid og jo større omkostninger vil der være forbundet med at opnå denne viden. Og hvis den ønskede viden yderligere skal have en høj grad af validitet, vil det typisk kræve en indsnævring af undersøgelsesfeltet, idet støjen skal holdes ude. Derved bliver det også sværere at anvende denne mere sikre viden i andre sammenhænge. Der skal ganske naturligt være et rimeligt forhold mellem indsats og omkostninger til at vurdere indsatsen.

5. Udviklingen i evalueringsteori

Tilgangen til evaluering af arbejdsmiljøindsatser er altså langt fra simpel og indlysende, når vi betragter evalueringsbehov ud fra de særlige karakteristika ved arbejdsmiljø som genstandsfelt. Det er derfor spørgsmålet, hvad evalueringsteorien og den tilknyttede værktøjskasse kan tilbyde. I dette afsnit redegør vi kort for udviklingen af evalueringsteori med særligt fokus på de aktuelle diskussioner om specielt evidensbegrebet, og vi slutter af med en kort diskussion af betydningen for evaluering af arbejdsmiljøindsatser.

Udviklingen kan sammenfattes i fire evalueringsbølger, som hver indeholder forståelser af hvad evaluering skal bruges til og hvordan det kan gennemføres (Krogstrup 2011):

- Den klassiske effektevaluering
- Responsiv evaluering
- Monitorering og evaluering
- Evidensbaseret evaluering

Som altid med denne type opdelinger er der ikke altid skarpe skillelinjer. Der er overlap, og nye tendenser bygger typisk videre på de foregående, samtidig med at disse typisk lever i bedste velgående. Samlet set giver de fire bølger dog et udmærket billede af hvordan udviklingen er forløbet op til i dag, og vi præsenterer de fire tilgange nedenfor primært med afsæt i Krogstrup (2011) (se også www.bar-eva.dk).

5.1. Den klassiske effektevaluering

Evaluering som begreb og aktivitet opstod i 1960'erne først i USA med en hurtig spredning til Europa. På det tidspunkt herskede der en forestilling om en konstant udvikling mod et bedre og rigere samfund, hvor det handlede om at finde frem til den rigtige udvikling. I USA blev det formuleret som "the experimental society", hvor systematiske forsøg kunne føre til afprøvning af forskellige politikker og derved kunne man finde frem til den rigtige måde at gøre tingene på. Der blev derfor udviklet et evalueringsdesign, som handlede om at måle situationen før en bestemt intervention, gennemføre interventionen og efterfølgende gennemføre en udfaldsmåling for på den måde at finde frem til effekten af interventionen. Dette design kunne forfines ved at anvende en kontrolgruppe, hvor der ikke gennemføres intervention, og effektmålingen kunne således styrkes ved både at have

før- og eftermålinger for interventionsgruppen og for kontrolgruppen. Denne type evaluering bygger stort set udelukkende på kvantitative dataindsamlings- og analysemetoder. Et væsentligt element var også at evaluator (eller forsker) skulle forholde sig objektivt til evalueringen hvilket bl.a. indebærer at evaluator skulle afholde sig fra at påvirke interventionen på nogen måde.

Dette evalueringsdesign er i høj grad stadig kendt og anvendt både bredt indenfor evaluering og indenfor arbejdsmiljøområdet i både evaluerings- og forskningsmæssig sammenhæng, og det er siden forfinet og ligger nu til grund for en væsentlig del af forståelsen af evidensbegrebet.

Det er imidlertid også et begreb som allerede på et tidligt tidspunkt blev kritiseret og fortsat kritiseres for en række forhold hvoraf særligt tre er væsentlige:

- Først og fremmest fungerede tankegangen om det vellykkede eksperiment som udgangspunkt for praktiske politiske programmer ikke. Evalueringerne viste sjældent særligt stor effekt, og politikkerne gennemførte under alle omstændigheder de programmer, som de af politiske årsager anså for rigtige.
- Rent metodisk blev denne tilgang desuden kritiseret for, at det ikke er muligt at udskille enkelt-elementer i samfundsmæssige interventioner, som kan undersøges adskilt fra, hvad der ellers sker i samfundet. Der vil altid ske en påvirkning fra andre samfundsmæssige udviklinger som kommer til at påvirke eftermålingen.
- Et sidste kritikpunkt er vanskelighederne med at etablere kontrolgrupper, som kun adskiller sig i forhold til den gennemførte intervention. Samfundsgrupper er aldrig ens og dermed påvirker interventioner grupper forskelligt.

5.2. Responsiv evaluering

Denne form for evaluering hoppede så at sige over i den modsatte grøft. Her begyndte man fra midten af 1970'erne at fokusere på faktiske programmer ude i samfundet som politikkerne under alle omstændigheder etablerer, og formålet med evaluering blev derfor set som forsøg på at bidrage til at gøre programmer så effektive som muligt. Det blev ikke anset som muligt at udskille egentlige slut-effekter af et program på grund af de mange faktorer, som influerer programmet. Ud fra den forståelse bliver det væsentligste at give feedback til de interessenter, som beskæftiger sig med programmet således, at de får muligheder for at forbedre implementeringen. Denne form for evaluering kaldes også procesevaluering eller formativ evaluering.

Den responsive evaluering betjener sig først og fremmest af kvalitative metoder i form af interview og forskellige former for observation. Kvantitative metoder anvendes dog også. Fx kan det i denne sammenhæng være relevant at anvende spørgeskemaer til at finde ud af om deltagere i et program har hørt om programmet og er blevet inddraget. Sådanne metoder indebærer at evaluator ikke kan forholde sig adskilt og objektivt fra programmet. Tværtimod bliver evaluator netop inddraget i programmet for derigennem at kunne forstå, hvad der foregår og efterfølgende give feedback til deltagerne, således at de kan forbedre deres indsats. Der er forskellige holdninger til, hvorledes evaluators inddragelse skal håndteres. I den mest oprindelige version skal evaluator godt nok være til stede i feltet for derved at observere, hvad der foregår, men uden i øvrigt at blande sig, og opgaven består i at videregive observationerne til interessenterne som derefter selv fortolker og beslutter sig til hvilke konsekvenser disse måtte have. I senere retninger inddrages evaluator i forskelligt omfang og sammenhæng mere aktivt i feedback mekanismerne. Der kan således være tale om at facilitere at

interessenterne selv i høj grad står for evalueringen fx ved at organisere dialogaktiviteter eller at træde ind i forskellige andre varianter af faciliterende eller coachende roller. Indenfor forskningen er denne rolle bl.a. udtrykt i aktionsforskning (Gustavsen 1992;Nielsen & Svensson 2006), som lægger op til, at felt og forsker i fællesskab ændrer virkeligheden og derigennem opnår både en ny og mere ønskværdig virkelighed og en forståelse af, hvordan denne virkelighed kan påvirkes.

Også den responsive evaluering er blevet udsat for kritik på flere punkter:

- Set fra den klassiske effektevaluering er den responsive evaluering uvidenskabelig, idet det bliver tilfældigt, hvad der fokuseres på og fravalget af effektmåling betyder at forsøgene på at optimere ikke har et grundlag af viden om effekt at basere sig på.
- Ud fra sine egne forudsætninger kan det desuden være problematisk, at evalueringen i for høj grad styres af evaluators og/eller interessenters egne præferencer og evalueringen kan dermed bl.a. få en konserverende effekt fordi der ikke stilles spørgsmålstegn ved programmet som sådan og optimeringen sker indenfor de givne rammer.

5.3. Monitorering og evaluering

I takt med fremvæksten af New Public Management (NPM) i 1980'erne vokser også den næste evalueringsbølge frem. NPM henter inspiration fra den private sektor og lægger vægt på markedsgørelse og målstyring. Det bliver derfor væsentligt at finde ud af om offentlige institutioner opfylder deres mål, om borgerne får de ydelser, de har krav på og om leverandører leverer det, de har lovet. Det bliver derfor væsentligt at finde parametre som kan belyse, om målopfyldelsen sker. Typisk skrives målene ind i resultatkontrakter for de enkelte institutioner og lederes løn og institutionerne budgetter bliver afhængige af graden af målopfyldelse. I de seneste års version er man i stigende grad igen inspireret af den offentlige sektor begyndt at anvende *Key Performance Indicators* (KPI) som omdrejningspunkt for den løbende monitorering.

Paradoksalt nok støder monitoreringen ind i den barriere for effektevaluering som vi beskrev under den klassiske effektevaluering og som også er en del af årsagen til, at det er vanskeligt at evaluere arbejdsmiljøindsatser. Der er meget langt fra de konkrete aktiviteter til de slutmål, som offentlige institutioner skal levere: fx skoleelever som fortsætter i en videregående uddannelse eller reduktion i antallet af førtidspensionister. Derfor kommer monitoreringen til at fokusere på procesvariable fx andelen af førtidspensionister som har været til samtaler eller deltaget i arbejdsprøvning, mens selve den ønskede effekt ikke bliver målt. En anden version er at lave brugertilfredshedsmålinger som kan være ganske vanskelige at knytte til de egentlige mål for institutionen. Fx viser patienttilfredshedsmålinger typisk høj tilfredshed (<http://patientoplevelser.dk/index.asp?id=499>), men spørgsmålet er, om denne tilfredshed udtrykker mere end patienternes oplevelse af, at personalet de møder optræder sympatisk. Det bliver simpelthen for dyrt og for langsigtet i forhold til både det daglige arbejde og de årlige budgettildelinger, hvis målstyringen skulle forholde sig til de langsigtede effektmål.

Der er derfor også peget på (Krogstrup 2011), at denne monitorering ikke har meget med evaluering at gøre, men som med de øvrige evalueringsbølger er den udtryk for et ønske om at lave systematisk vidensindsamling, som gennem en bedømmelse kan anvendes som fremtidigt beslutningsgrundlag. Det er samtidig en tilgang til evaluering som i høj grad danner grundlag for nutidens dominerende evalueringsbølge.

5.4. Evidensbaseret evaluering

Efter årtusindskiftet er evidensbaseret evaluering i særlig grad sat i højsædet. Et krav om at offentlige midler skal anvendes, hvor der er evidens for effekten er på mange måder en naturlig forlængelse af NPM og monitoreringsbølgen. Det gælder ikke mindst indenfor arbejdsmiljøet hvor den tætte tilknytning til det biomedicinske forskningsparadigme skaber en særlig grobund for kravet om evidens for arbejdsmiljøindsatser. Der er imidlertid ikke enighed om, hvordan evidens skal forstås, og der udspiller sig ofte heftige diskussioner om, hvornår der er tilstrækkelig evidens for at iværksætte særlige tiltag. Det er også en velkendt diskussion fra arbejdsmiljøområder, hvor der ofte er diskussioner af evidensen for, at noget er skadeligt for helbredet og om bestemte forebyggende foranstaltninger har en effekt eller ej. Der er særligt to positioner som dominerer debatten. Den ene bygger på den klassiske effektevaluering og inspireret af USA er det randomiserede kontrollerede forsøg (randomized controlled trials – RCT) blevet udpeget som den gyldne standard. Den anden position bygger på kritisk evaluering (Pawson 2006) og fremhæver viden om mekanismer og kontekst som afgørende for evidens.

Effektevaluering

RCT-tilgangen henter i høj grad inspirationen fra USA og bygger på den medicinske forskning, hvor effekten af medicin undersøges gennem sammenligning af den aktive ingrediens og kalkpiller (placebo). Der trækkes lod mellem deltagerne, og forskerne ved ikke, hvem der får hvilken type piller. Ideen med denne metode er at sikre fuldstændig sammenlignelighed, således at den eneste forskel er medicinen (interventionen). I denne evalueringstilgang overføres denne metode til andre anvendelsesområder som både omfatter individer og organisationer. Der kan fx trækkes lod mellem belastede unge om deltagelse eller ikke deltagelse i afprøvning af støtteformer. Med hensyn til organisationer kan man fx trække lod mellem skoler eller klasser om afprøvning af bestemte undervisningsmaterialer. På arbejdsmiljøområdet er metoden først og fremmest anvendt til interventioner, som retter sig mod enkeltindivider. Et typisk eksempel kan være tilbud om træning som afgøres ved lodtrækning, og hvor man eventuelt efterfølgende giver kontrolgruppen det samme tilbud. Man kan også tænke sig randomisering på organisationsniveau, hvor man fx trækker lod mellem teams eller mindre enheder om deltagelse/ikke deltagelse. Metoden kan dog ikke anvendes i den ideelle form, fordi de deltagere som gennem lodtrækningen ikke får interventionen, er bekendt hermed og dermed på forskellig måde kan blive påvirket af situationen. Netop fordi det kan være svært at randomisere, anerkendes også metoder med anvendelse af kontrolgrupper og før-og-efter-målinger som relevante om end metodemæssigt svagere muligheder. I denne tilgang anvendes i høj grad systematiske reviews af litteraturen, hvor de anvendte forskningsmetoders kvalitet bedømmes, og der er udarbejdet internationale modeller for, hvorledes det skal gennemføres i praksis. Den medicinske model betegnes Cochrane (<http://www.cochrane.org>) og den samfundsvidenskabelige model Campbell (<http://www.campbellcollaboration.org>).

Kritisk evaluering

Kritisk evaluering som i Danmark ofte kaldes virkningsevaluering, er i særlig grad udviklet indenfor de seneste år (Dahler-Larsen & Krogstrup 2004; Pawson 2006; Pawson & Tilley 1997; Pawson, Wong, & Owen 2011). Udgangspunktet i denne forståelse er, at der indenfor samfundsmæssige interventioner yderst sjældent findes sikre effekter som er kontekstuafhængige. Det betyder, ifølge den kritiske evaluering, at selv den bedst gennemførte effektevaluering med anvendelse af RCT-

lignende metoder ikke kan udtale sig om effekten udenfor den kontekst den foregår i. Det væsentlige består derfor i at finde frem til de mekanismer, som i samspil med konteksten kan føre til et bestemt resultat. Dermed bliver det muligt at finde ud af, hvad der virker og hvordan det virker, og gennem forståelsen af mekanismerne kan denne viden overføres til andre sammenhænge. Den kritiske evaluering benytter sig af både kvalitative og kvantitative metoder. På samme måde som effekt-evalueringen gerne benytter sig af review, anbefaler den kritiske evaluering også, at der udarbejdes systematiske reviews. Det skyldes et behov for at have flere evalueringer af sammenlignelige indsatser, for at kunne få en bedre forståelse af mekanismerne, idet en enkelt evaluering ikke vil være tilstrækkeligt omfattende til at kunne tegne et generelt billede af mekanismerne. Mens effektevalueringen har en mangeårig tradition bag, er den kritiske evaluering endnu relativt ung, og erfaringerne med at anvende metoden er under opbygning. Indenfor arbejdsmiljøområdet er der endnu kun spredte eksempler på denne type evaluering (Se fx Pedersen, Nielsen, & Kines 2012). En af vanskelighederne som den kritiske evaluering skal overvinde – ikke mindst i arbejdsmiljøsammenhæng, er at få en mere præcis beskrivelse af, hvad en mekanisme er og hvordan den fungerer. Et af midlerne hertil er en systematisk anvendelse af programteori, hvor interessenterne på forhånd opstiller en teori for, hvordan interventionen vil virke. Holdbarheden af denne programteori kan undersøges i evalueringen, idet man får et grundlag for at undersøge mekanismer, da programteorien på forhånd har opstillet hypoteser for de mekanismer, som skal få interventionen til at virke.

Pragmatisk evaluering

I den praktiske verden foretages kun få evalueringer som klart følger den ene eller den anden retnings metodekrav. Når det sker, drejer det sig typisk om egentlig forskning, som stiller større krav til systematisk metodeanvendelse. De fleste evalueringer plukker fra hylderne og indeholder både kvantitative og kvalitative elementer. Kvantitative effektmål findes måske i den løbende monitoring eller brugere kan spørges om deres holdning, og der gennemføres kvalitative interview af, hvad der foregår i praksis og som dermed måske kan sige noget om mekanismer som er i spil, og ofte gennem midtvejsevalueringer forsøger at give feedback for at kunne styrke implementeringen af interventionen. Årsagen er udpræget pragmatisk. Det er begrænset hvor mange ressourcer, der kan anvendes til evaluering, og der er typisk mange interesser som evalueringen skal tilgodese. Indvendingerne mod den pragmatiske fremgangsmåde vil være, at den sætter sig i mellem to stole. Når de kvantitative effektmål ikke er fremkommet gennem RCT eller anvendelse af systematiske kontrolgrupper, ved man i virkeligheden ikke om effekten skyldes interventionen eller noget andet. Og tilsvarende vil begrænsede interview af de involverede partner måske give et billede af deres oplevelser, men det vil ikke kunne svare på hvilke mekanismer, som faktisk har drevet processen igennem.

6. Evidens og arbejdsmiljøindsatser

Vi slutter af med overvejelser om forståelsen af evidensbegrebet og hvordan det kan kobles til evaluering af arbejdsmiljøindsatser. I den første halvdel af dette notat beskrev vi, hvorledes arbejdsmiljøindsatsen har bevæget sig fra at fokusere på konkrete arbejdsmiljøproblemer med konkrete løsninger til komplekse problemer som fx stress og sygefravær som også har komplekse løsninger. I evalueringsteorien er det også blevet beskrevet som 'tamme' og 'vilde' problemer. De tamme problemer kan løses af en entydig intervention med en sikker effekt. Det gælder eksemplet med af-

skærmning af rundsaven. De vilde problemer er derimod komplekse. De kan vanskeligt beskrives og afgrænses, og effekten er stærkt kontekstafhængig og mere eller mindre uforudsigelig. I bedste fald kan der derfor tales om en sandsynlig effekt af en intervention. Selvom det er muligt at fremskaffe mere og mere viden om de vilde problemer og om mulige interventioner som kan bidrage til at øge sandsynligheden for en positiv effekt, vil de aldrig kunne reduceres til tamme problemer. De stadige forandringer af konteksten vil betyde at udfaldet vil blive mere eller mindre forskelligt hver gang.

Denne problemstilling betyder imidlertid ikke at effektevaluering skal forkastes, da arbejdsmiljøet i dag i høj grad vil være domineret af vilde problemer. Der er således behov for at skaffe viden om, i hvilken grad interventioner medfører kvantificerbare ændringer. Hvis der ikke er tegn på ændringer, vil der næppe være belæg for at fortsætte med en bestemt indsats, selvom denne indsats også er stærkt kontekstafhængig. Det er blot vigtigt at være opmærksom på, at en positiv effekt godt kan optræde, selvom et niveau er uændret eller negativ. Det kan være, at ændringer i konteksten medfører væsentlige forværringer som indsatsen faktisk forebygger en væsentlig del af – uden dog at kunne opnå en positiv udvikling. Det var faktisk tilfældet i et projekt, der undersøgte sygefravær i københavnske plejehjem. Her steg sygefraværet betydeligt i løbet af projektperioden på alle plejehjem, men i interventionsplejehjemmene var sygefraværet signifikant mindre (Kristensen & Smith-Hansen 2003). Der skal desuden tages hensyn til at effekt på arbejdsmiljø og helbred typisk optræder som resultat af mange forskellige indsatser, som spiller sammen. Det vil således sjældent være muligt at finde sådanne direkte effekter af enkeltstående indsatser.

Erkendelsen af at de fleste arbejdsmiljøindsatser optræder som vilde problemer, hvor der ikke kan forudsiges en sikker effekt, betyder naturligvis ikke, at der ingenting skal gøres. Vi ved, der findes væsentlige arbejdsmiljøproblemer, som ganske vist er vanskelige at løse, men som har alvorlige konsekvenser for både individer, virksomheder og samfund. Det er derfor under alle omstændigheder nødvendigt at handle, og en for snæver forståelse af evidensbegrebet kan hermed hæmme en nødvendig udvikling af nye bedre indsatser. Evidens som begreb må derfor forstås som den bedst tilgængelige viden om hvordan indsatser skal designes for at opnå den størst mulige sandsynlig for en høj grad af effekt. Den tilgængelige viden skal anvendes her og nu, samtidig med, at der arbejdes så systematisk som muligt med at skaffe yderligere viden, og heri spiller fortsatte evalueringer en væsentlig rolle i forøgelsen af vores vidensgrundlag.

Litteratur

Advice A/S 2008, *Evaluering af videncenter for arbejdsmiljø – Centerets virksomhed 2005 til primo 2008*, Videnscenter for Arbejdsmiljø.

Arbejdsmiljøgruppen af 1972 1973, *Arbejdsmiljø: skader, omkostninger, målsætninger, nærdemokrati, planlægning*, Arbejdsmiljøgruppen, 1.

Arbejdsmiljøgruppen af 1972 1974, *Arbejdsmiljø: dårlige rygge, stress, høreskader*, Arbejdsmiljøgruppen, 2.

Arbejdsmiljøgruppen af 1972 1975, *Arbejdsmiljø: grænseværdier, arbejdstiden, meningsfyldt job, udstødningen*, Arbejdsmiljøgruppen, 3.

Arbejdsmiljøgruppen af 1972 1976, *Arbejdsmiljø: medbestemmelse, arbejdsmiljøtjeneste, uddannelse og forskning, planlægning*, Arbejdsmiljøgruppen, 4.

Arbejdstilsynet 2007, *10 år med Rent Arbejdsmiljø 2005*, Arbejdstilsynet.

Arbejdstilsynet 2010, *Evaluering af arbejdsmiljøreform - Slutrapport*, Arbejdstilsynet.

Arbejdstilsynet 2011, *Evaluering af Arbejdsmiljøfonden*, Arbejdstilsynet.

Black, J. & Baldwin, R. 2010, "Really Responsive Risk-Based Regulation", *Law & Policy*, vol. 32, no. 2, pp. 181-213.

Briner, R. B. & Reynolds, S. 1999, "The costs, benefits, and limitations of organizational level stress interventions", *Journal of Organizational Behavior*, vol. 20, no. 5, pp. 647-664.

Capacent & Arbejdstilsynet 2007, *Arbejdstilsynet Effektevaluering af BAR-projekter*, Arbejdstilsynet.

Cox, A., O'Reaganm, S., Denvir, A., Broughton, A., Pearman, D., Tyers, C., & Hillage, J. 2008, *What works in delivering improved health and safety outcome – A review of the existing evidence*.

Dahler-Larsen, P. 2010, "Evaluering," in *Kvalitative metoder - en grundbog*, S. Brinkman & L. Tanggaard, eds., Hans Reitzels Forlag, København, p. 165.

Dahler-Larsen, P. & Krogstrup, H. K. 2004, *Nye veje i evaluering* Systime Academic.

Frick, K. & Wren, J. 2000, "Reviewing Occupational Safety and Health Management," in *Systematic Occupational Health and Safety Management*, K. Frick, P. L. Jensen, & T. Wilthagen, eds., Pergamon, Oxford, pp. 17-42.

Gustavsen, B. 1992, *Dialogue and development : theory of communication, action research and the restructuring of working life* Van Gorcum.

Hansen, A. M., Starheim, L., & Nielsen, K. T. 2009, *Tilsyn, regulering og virkning på arbejdsmiljøområdet – et litteraturreview med særligt focus på psykisk arbejdsmiljø*, RUC, Roskilde.

Hasle, P. & Zwetsloot, G. 2011, "Editorial: Occupational Health and Safety Management Systems: Issues and challenges", *Safety Science*, vol. 49, no. 7, pp. 961-963.

Health and Safety Executive 2012, *Health and Safety Executive - Business Plan 2010/2011*, Health and Safety Executive.

Kristensen, T. S. & Smith-Hansen, L. 2003, *Det udviklende arbejde: Helbred, stress og kvalifikationer. Resultater fra SARA-projektet* Frydenlund, København.

- Krogstrup, H. K. 2011, *Kampen om evidens - resultatmåling, effektevaluering og evidens* Hans Reitzels Forlag, København.
- Møller, N., Jensen, P. L., & Broberg, O. 1988, *Arbejds måder i sikkerhedsgruppen*, Arbejds miljø-fondet, København.
- Nielsen, K. A. & Svensson, L. 2006, *Action Research and Interactive Research* Shaker Publishers, Maastricht.
- Nielsen, K. T., Meldgaard Hansen, A., Lund, H. L., & Starheim, L. 2010, *Har Arbejdstilsynet et godt øje til arbejdsmiljøet*, Landsorganisationen i Danmark.
- Pawson, R. 2006, *Evidence-based policy: A realist perspective* Sage.
- Pawson, R. & Tilley, N. 1997, *Realistic evaluation* Sage, Los Angeles, London, New Delhi, Singapore, Washington DC.
- Pawson, R., Wong, G., & Owen, L. 2011, "Known Knowns, Known Unknowns, Unknown Unknowns: The Predicament of Evidence-Based Policy", *American Journal of Evaluation*, vol. 32, no. 4, pp. 518-546.
- Pedersen, L. M., Nielsen, K. J., & Kines, P. 2012, "Realistic evaluation as a new way to design and evaluate occupational safety interventions", *Safety Science*, vol. 50, no. 1, pp. 48-54.
- Rigsrevisionen 2000, *Arbejdstilsynets styring af tilsynsfunktionen*, Rigsrevisionen.
- Robson, L., Clarke, J., Cullen, K., Bielecky, A., Severin, C., Bigelow, P., Irvin, E., Culyer, A., & Mahood, Q. 2007, "The Effectiveness of Occupational Health and Safety Management Systems: A Systematic Review", *Safety Science*, vol. 45, pp. 329-353.
- Semmer, N. K. 2006, "Job stress interventions and the organization of work", *Scandinavian Journal of Work Environment & Health*, vol. 32, no. 6, pp. 515-527.
- Stranddorf, J., Møller, N., & Jensen, P. L. 1992, *Arbejds miljøarbejdet i udvikling*, Arbejds miljø-fondet, København.
- Vedung, E. 1998, "Policy instruments: typologies and theories," in *Carrots, sticks & sermons - policy instruments and their evaluation*, M.-L. Bemelmans-Videc, R. C. Rist, & E. Vedung, eds., Transaction Publishers, New Brunswick, NJ, pp. 21-58.